A transdiagnostic, natural language processing analysis of mental disorder symptom burden using electronic health record data

Rashmi Patel[1,2], Soon Nan Wee[1], Jesisca Tandi[1] and Joydeep Sarkar[1]
[1]Holmusk Technologies, Inc., New York, USA., [2]King's College London, UK

BACKGROUND

Criterion-based diagnostic classification does not necessarily capture the breadth of clinical presentation in mental disorders. A transdiagnostic, symptom-based approach may enable a better classification of clinical phenotype at the individual patient level.[1]

Psychiatrists document clinical information in electronic health records (EHR). Natural language processing (NLP) tools may be used to extract mental state examination (MSE) from unstructured free text in EHRs.

OBJECTIVE

We applied NLP to a large de-identified EHR dataset to investigate the transdiagnostic distribution of mental disorder symptom data.

METHODS

- Data from 543,849 patients (Table 1) were obtained from NeuroBlu, a secure trusted research environment (TRE)[2] that enables the analysis of de-identified, HIPAA-compliant data from 25 U.S. healthcare providers that employ the MindLinc EHR system.

- NLP models[3] were used to extract the following MSE features:
(i) Delusions/hallucinations: Delusions of grandeur; persecutory delusions; paranoid delusions; hyper-religiosity; auditory, visual, tactile and olfactory hallucinations
(ii) Mood: Anxious, depressed; irritable; elevated; labile; suicidal
(iii) Cognition: Impaired; poor attention; poor concentration

- These features were chosen because they represent key domains in the psychopathology of mental disorders

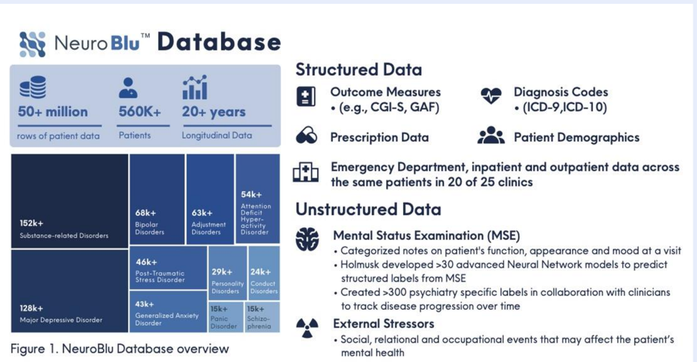

Figure 1. NeuroBlu Database overview

Data Source of US Health Facilities

De-identified EHR data were obtained from U.S. mental health services that use the MindLinc EHR system. The data were analyzed in NeuroBlu, a secure Trusted Research Environment (TRE) that enables data assembly and analysis using an R/Python code engine.[2]

RESULTS

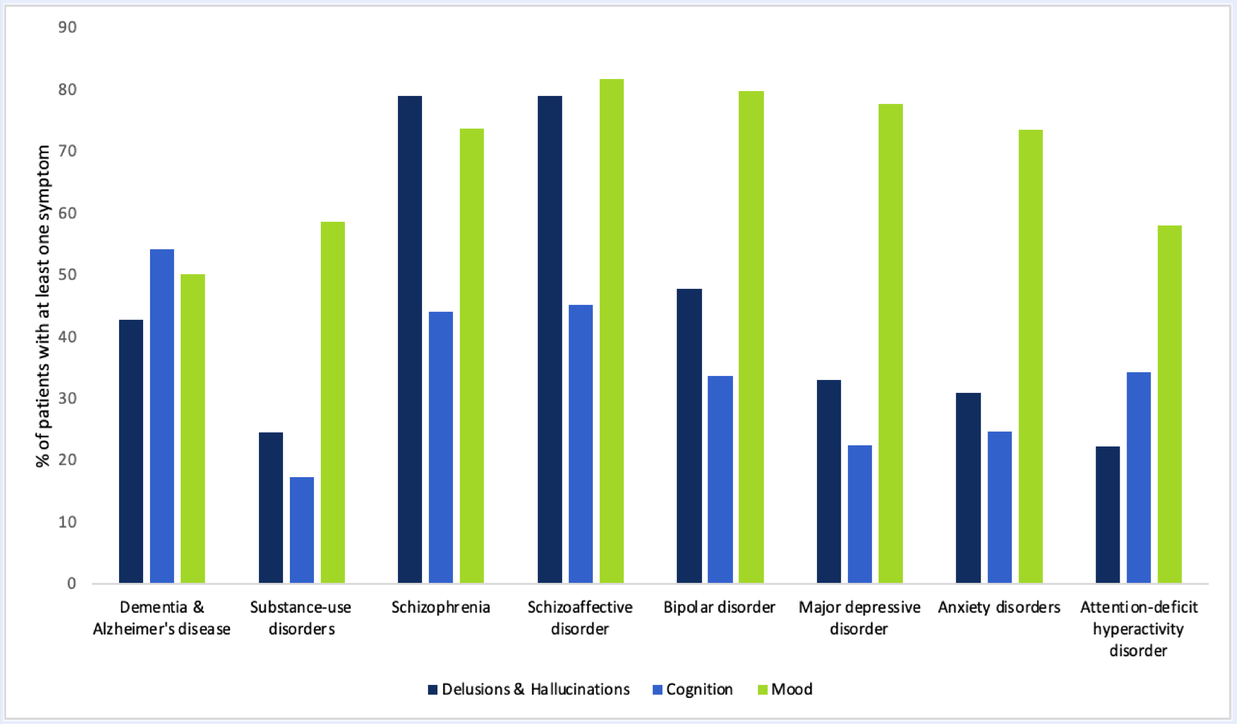

Figure 2. Prevalence of at least one NLP-derived Mental State Examination symptom (%) according to diagnostic category. Information on comorbidities was not extracted for this analysis.

Key Findings

- Delusions/hallucinations were most prevalent in schizophrenia and schizoaffective disorder.

- Cognitive problems were most prevalent in dementia/Alzheimer's disease but were also present in psychotic disorders.

- Mood symptoms were often prevalent among diagnostic categories which are not typically considered to be a mood disorder (e.g., schizophrenia).

Diagnosis	N
Dementia/Alzheimer's disease	5,029
Substance use-disorder	152,790
Schizophrenia	14,592
Schizoaffective disorder	15,044
Bipolar disorder	69,607
Major depressive disorder	129,120
Anxiety disorder	103,923
Attention-deficit hyperactivity disorder	52,744

Table 1. Sample sizes for diagnostic categories within the NeuroBlu database.

CONCLUSION

- These findings illustrate the potential for a transdiagnostic approach to better represent clinical phenotype than criterion-based diagnosis alone.

- These findings also illustrate the potential for NLP-derived measures to support clinical decision making using EHR data.

Conflicts of Interest: All authors report personal fees from Holmusk Technologies, Inc. RP reports equity ownership in Holmusk Technologies, Inc.

References:

1. Fusar-Poli P, Solmi M, Brondino N, et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry* 2019;**18**:192-207.
2. Patel R, Wee S N, Ramaswamy R, Thadani S, Ruchir G, Calvanese N., et al. Cohort profile: NeuroBlu, an electronic health record trusted research environment to support mental healthcare analytics with real-world data. BMJ Open, Accepted/In Press.
3. Mukherjee SS, Yu J, Won Y, et al. Natural Language Processing-Based Quantication of the Mental State of Psychiatric Patients. *Comput Psychiary* 2020;**4**:76.